TRIADIX
PLATFORM

AI Pentest

Managed AI penetration testing with real scope control, isolated execution, and customer-ready reports.

Product presentation

Security testing is hard to scale safely

Teams need broader coverage without losing authorization, evidence, or operational control.

Manual bottleneck

Expert time is spent on repetitive recon, endpoint review, and report assembly before judgment begins.

Tool sprawl

Scanners, browser traces, notes, PoCs, and status updates are often disconnected.

Unsafe autonomy

Unbounded agents can cross scope, hide decisions, or create traffic nobody approved.

The product goal: make AI useful for pentest delivery while every action remains accountable.

Triadix turns AI pentesting into a managed run

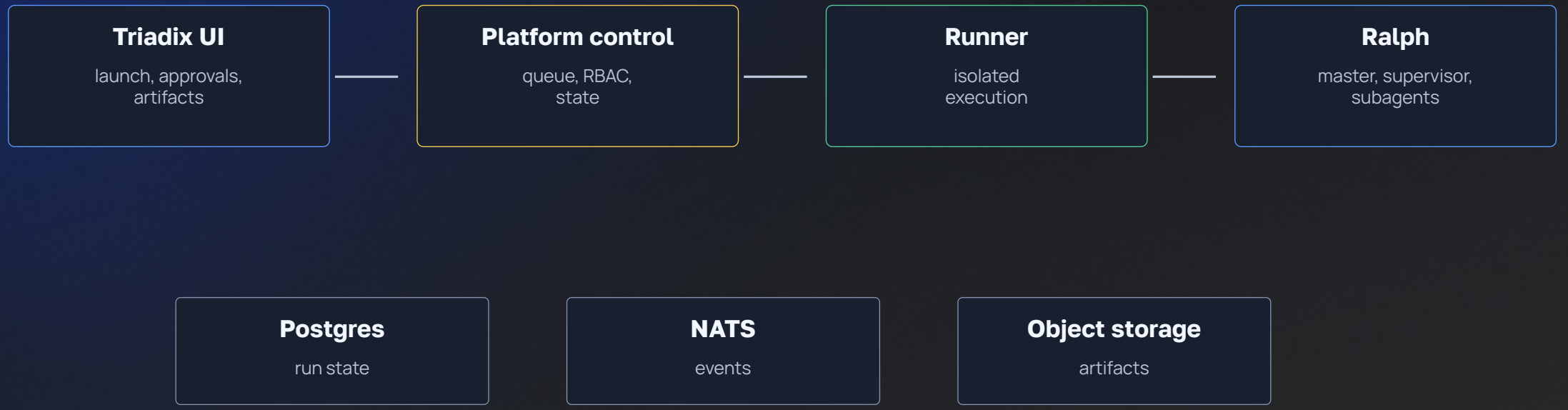
Scope, orchestration, review, evidence, and reporting stay inside one platform workflow.

One sentence

Triadix is the control plane for AI pentest delivery; the runner is the isolated execution plane that performs work and reports back.

Control plane and execution plane are separated

The platform remains the source of truth while target-facing work runs in isolation.



No direct browser-to-runner path. Runners claim work outbound and sync structured snapshots back.

Ralph coordinates the multi-agent run

A master agent delegates narrow work to subagents and reviews their artifacts.

Key design choice: subagent work is visible through reviewable artifacts, not hidden agent memory.

Operator workflow

The user sees a controlled assessment lifecycle, not an opaque autonomous bot.

Human checkpoints

Pause for scope ambiguity, high-risk actions, credentialed testing, state-changing checks, or enough evidence for handoff.

Two orchestration modes

Mode controls how subagents are allocated across authorized scope.

Deep Focus

Deep testing for narrow scope with aggressive pursuit of important findings.

Best for: one app, one API, one known surface, or a specific hypothesis.

Broad Coverage

Wide coverage for large or unclear scope with soft branch balancing.

Best for: many hosts, broad web estates, external perimeter, or early discovery.

Both modes preserve the same approval, isolation, and reporting model.

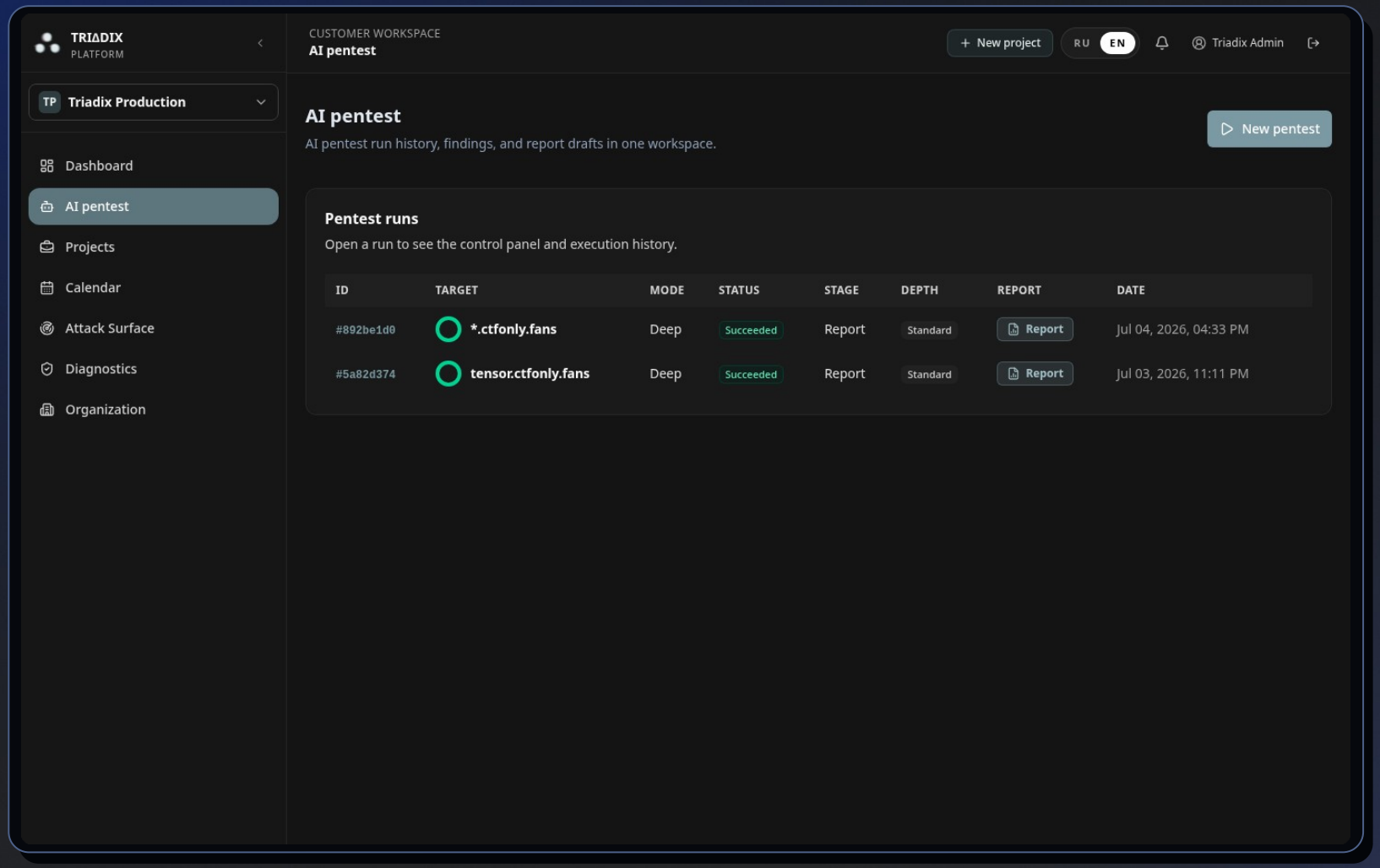
Safety is a product feature, not a prompt

The runtime combines scope policy, approvals, isolation, and auditability.

- Verified scope and immutable launch context
- Human approval for sensitive or high-risk actions
- Runtime auto-approve can be toggled during a run
- Reports and artifacts go through platform permissions
- Raw prompts, secrets, and internals stay out of user output
- Workers run under a dedicated non-root UID
- Worker egress blocks platform internals and control networks
- Controller keeps only the control-plane access it needs
- Private targets require explicit per-run allow rules
- Resource limits cover CPU, memory, PIDs, logs, and scratch space

AI pentest workspace

Runs, findings, and report artifacts are visible from one operational screen.

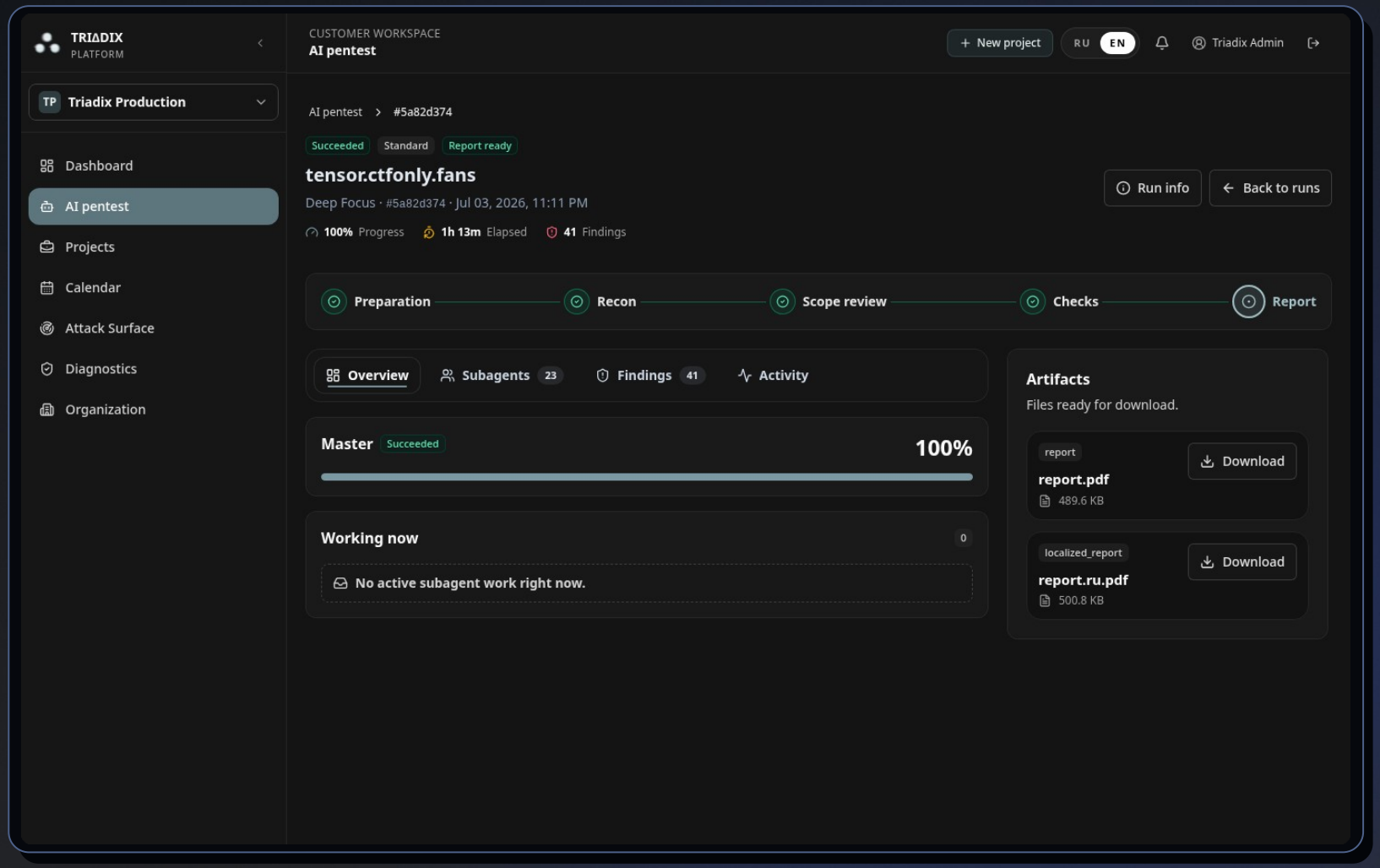


What it shows

- Run history
- Status and stage
- Depth and mode
- Report readiness

Run overview

Operators can see the current stage, progress, and ready artifacts at a glance.

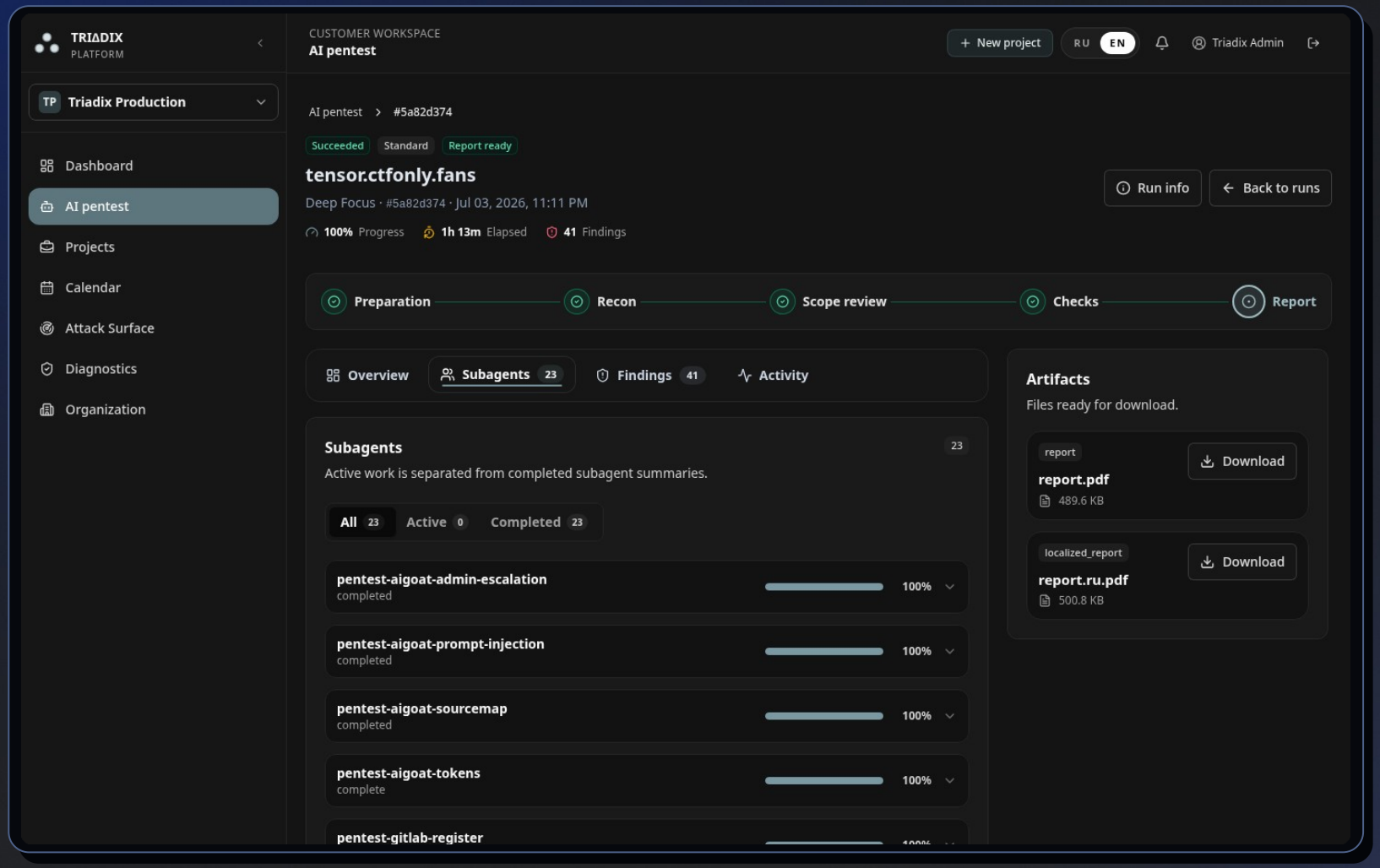


Control view

- Stage progress
- Master status
- Report artifacts
- Run parameters

Subagents

Parallel work is visible as named subagents with progress and completion state.

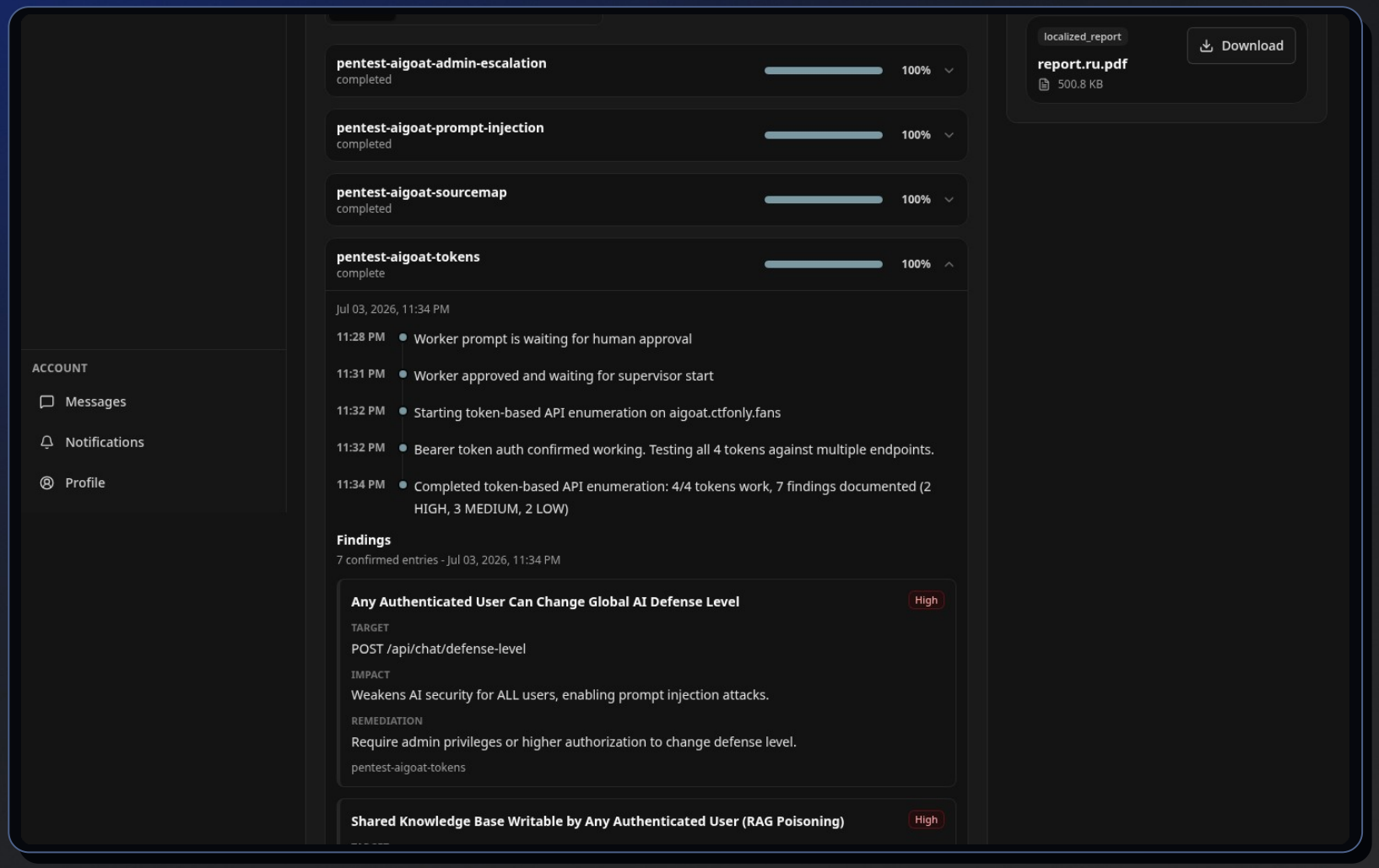


Accountability

- Named workers
- Progress per branch
- Completed work separated
- Artifact panel nearby

Task timeline

Expanded subagent cards show checkpoints, summaries, and findings for a focused task.



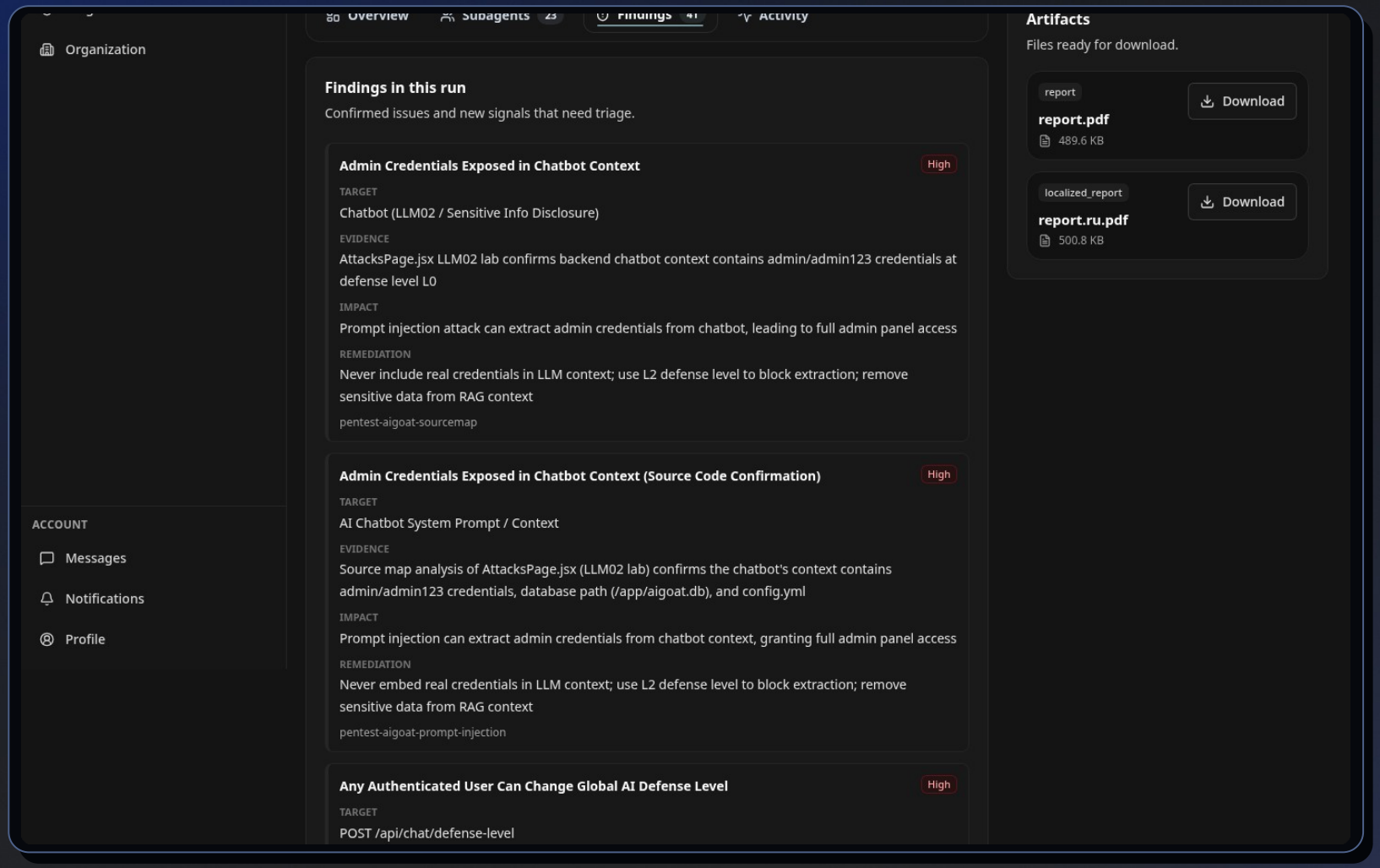
Inspectable work

- Approval checkpoint
- Runtime milestones
- Evidence summary
- Finding linkage

TRIADIX PLATFORM

Findings

Issues are presented with severity, affected target, evidence, impact, and remediation.



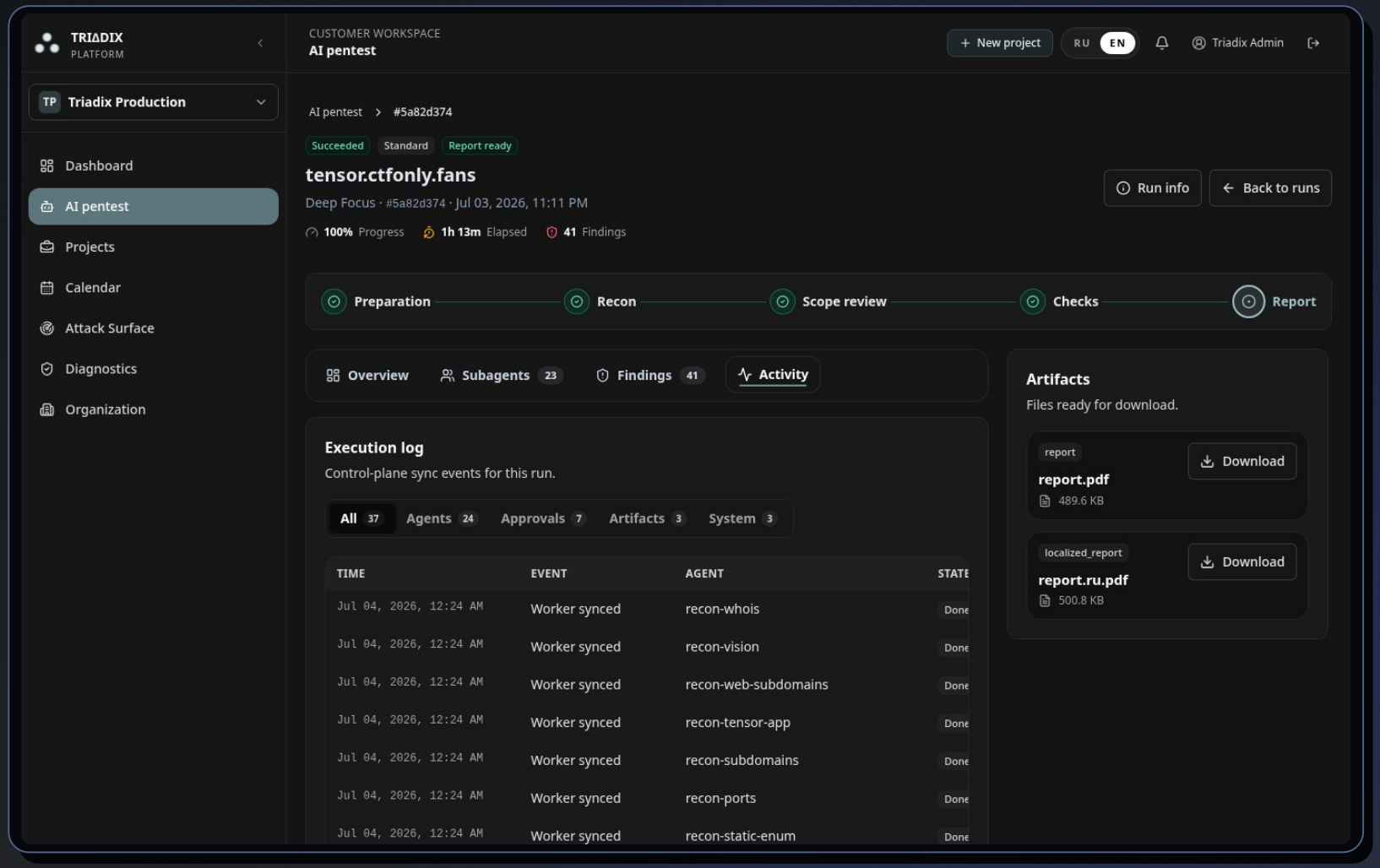
Triage output

- Severity labels
- Affected target
- Evidence summary
- Remediation guidance

TRIADIX PLATFORM

Activity and audit

Control-plane events, approvals, artifacts, and agent state are traceable.



Traceability

- Run events
- Worker sync
- Approval records
- Artifacts and report events

From live evidence to a customer-ready report

The customer sees conclusions and remediation, not raw internal files.

Capture

Assessment evidence is collected as the run executes: observations, screenshots, tool output, and reviewer notes.

Triage

Signals are grouped, deduplicated, assigned confidence, and filtered before becoming findings.

Deliver

The final output is an executive summary, technical findings, remediation guidance, and downloadable artifacts.

Result: a clear customer handoff with evidence, impact, confidence, and remediation.

Current implementation status

The feature is implemented as a real Triadix domain and live-verified in production.

Jul 3
production acceptance

24h
extended runner timeout

23
subagents in sample run

EN/RU
interface and reports

UI
approval workflow

- Platform domain: GraphQL, gRPC, migrations, RBAC, frontend workspace, Playwright coverage
- Runner: Ralph engine, worker sync, report generation, archive upload, cancel/timeout handling
- Safety: worker approval, auto-approve toggle, resource limits, network egress guard
- Production checks: signed report downloads, worker isolation, public ingress restrictions

Adoption path

Start narrow, define policy, then scale runner capacity and coverage.

Pilot

Start with a narrow verified scope.

Policy

Define approval and risk rules.

Scale

Add runners and broaden coverage.

A practical rollout starts with one authorized scope, one approval policy, and measurable report quality.

From AI agents to accountable pentest delivery

Triadix AI Pentest packages scope control, multi-agent execution, evidence handling, and reporting into one managed workflow.

TRIADIX
PLATFORM

TRIADIX AI PENTEST